Parameter Estimation and Model Testing for Markov Processes via Conditional Characteristic Functions

Song Xi Chen[a,b] [1], Liang Peng[c] and Cindy L. Yu[a]

[a] Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, USA

[b] Guanghua School of Management, Peking University, Beijing, 100871, China

[c] Department of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332

Abstract: Markov processes are used in a wide range of disciplines including finance. The transitional densities of these processes are often unknown. However, the conditional characteristic functions are more likely to be available especially for Lévy driven processes. We propose an empirical likelihood approach for estimation and model specification test based on the conditional characteristic function for processes whose sample paths can be either continuous or discontinuous with jumps. An empirical likelihood estimator for the parameter of a parametric process, and a smoothed empirical likelihood ratio test for the parametric specification of the process are proposed, which are shown to have good theoretical properties and empirical performance. Simulations and empirical case study are carried out to confirm the effectiveness of the estimator and the test.

Keyword: Conditional characteristic function; Diffusion processes; Empirical likelihood; Kernel smoothing; Lévy driven processes

[1]Corresponding Author: Song X. Chen, Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, USA. Tel: 1-515-2942729. Fax: 1-515-294 4040; Email: songchen@iastate.edu.

1 Introduction

Let $(X_t(\theta))_{t\in\mathcal{T}}$ be a parametric specification of a d-dimensional Markov process with a finite dimensional parameter $\theta \in \Theta \subset R^p$ where the time set $\mathcal{T}$ can be either a discrete set like $\{0, 1, \ldots\}$ or a continuous interval $[0, T]$, which corresponds to a discrete-time or a continuous-time Markov processes respectively. Despite the underlying process may be continuous in time, the process is observed only at discrete time points, say $0, \delta, \ldots, n\delta$. Here δ is the sampling interval. For a discrete time process, we assume $\{j\delta\}_{j=0}^n \subset \mathcal{T}$.

For the ease of illustration, we will concentrate on continuous-time Markov processes defined by stochastic differential equations. It is apparent from the rest of the paper that the proposed statistical inference framework based on the conditional characteristic function readily applies to discrete-time Markov processes. Specifically, the class of processes we will be focusing on is defined by

$$dX_t = \mu(X_t; \theta)dt + \sigma(X_t; \theta)dL_{t;\theta}, \tag{1.1}$$

where the state variable $X_t \in R^d$ consists of d-components, $\mu(\cdot)$ is a d-dimensional drift function, $\sigma(\cdot)$ is a $d \times d$ matrix-valued function of X_t, and L_t is a Lévy process in R^d. When L_t is a standard Brownian motion, (1.1) specifies a diffusion process which is known for having a continuous sample path. When L_t contains the Brownian motion and a compound Poisson process, (1.1) becomes the jump diffusion process. A stochastic process of form (1.1) has long been used to model dynamic stochastic systems arising in physics, biology and other natural sciences. It has been the fundamental tool in modern financial modeling and evaluation. See Sundaresan (2000) and Fan (2005) for overviews on financial applications, and Barndorff-Nielsen, Mikosch and Resnick (2001) for recent developments on Lévy driven processes. Statistical inference based on continuously sampled observations is discussed in Sørensen (1991).

Often a closed form expression for the transitional density of process (1.1), if it exists and is unique, is not available except for some special processes. This fact prevents the use of the maximum likelihood estimation method and the specification tests based on the exact transitional density. For diffusion processes, Aït-Sahalia(2008) and (2002) and Aït-Sahalia and Kimmel (2002) have established Edgeworth type expansion for the transitional densities, which facilitates approximate maximum likelihood estimation and tests. Since the density expansion is usually available for a limited number of terms, the sampling interval δ has to be small to allow the consistency in both estimation and testing.

Despite the transitional density of a process usually does not have a closed form expression, the conditional characteristic function (CCF) is more likely to be known explicitly. This is the case for affine jump diffusion processes (Duffie, Pan and Singleton, 2000) and time-changed Lévy processes (Chernov, Gallant, Ghysels and Tauchen, 1999). Also, characteristic functions are powerful tools in analyzing Lévy processes as demonstrated by the celebrated Lévy-Khintchine formula. Since an explicit CCF will allow exact estimation and testing for any given sampling frequency without the need to assume high data frequency, statistical inference based on the CCF can become more widely applicable and attractive. Recently, Aït-Sahalia and Yu (2006) established saddle point approximations to the CCF for a class of processes within (1.1). Our proposed approach can be applied to this approximated CCF provided the sampling interval is small.

Statistical inference based on the characteristic functions was proposed by Feuerverger and Mureika (1977), Feuerverger and McDunnough (1981) for independent observations and Feuerverger (1990) for discrete time series, which has received renewed interest recently. Singleton (2001), Chacko and Viceira (2003) and Carrasco, Chernov, Florens and Ghysels (2007) propose using the CCF for parameter estimation of multivariate processes through generalized method of moment (GMM) estimator based on either conditional moments implied from the CCF or a large number of estimating equations. Jiang and Knight

(2002) propose GMM estimators based on the joint characteristic function of the observed state variables. Chen and Hong (2005) propose a test for multivariate continuous-time models based on the CCF via a generalized spectral density approach.

In this paper, we propose an empirical likelihood (Owen, 1988) approach for parameter estimation and model specification testing of a parametric Markov process via the CCF. An empirical likelihood ratio is formulated for the unknown parameters assuming the parametric specification (1.1), which leads to a nonparametric likelihood parameter estimator. The estimator utilizes a wide range of frequency information in the parametric CCF, and yet the computation is easily managed without encountering high dimensionality. Our simulation study shows that the proposed parameter estimator can estimate model parameters accurately using discretely sampled data. A smoothed empirical likelihood ratio is proposed for testing the parametric model specification (1.1). We demonstrate by both theoretical and empirical studies that the proposed test procedure captures model misspecification within the model class (1.1). The tests considered in the paper include identifying a diffusion process and a finite-activity jump process.

The paper is organized as follows. In Section 2, we outline a range of processes within (1.1) and present their CCFs. Section 3 introduces and evaluates the CCF based empirical likelihood estimator. The model specification test is given in Section 4. Section 5 reports results from simulation studies. An empirical study for a set of 3-month treasury bill rate data is analyzed in Section 6. All technical details are reported in the Appendix.

2 Processes and Their CCFs

In this section, we outline some important members of Markov processes encompassed by (1.1). Let $\psi_t(u;\theta) = E(e^{iu^T X_{t+1}}|X_t)$, for $u \in R^d$, be the conditional characteristic function of the parametric process (1.1). We start with a multivariate diffusion process defined by

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dB_t, \tag{2.1}$$

where B_t is the standard Brownian motion in R^d. The existence and uniqueness of a process $\{X_t\}$ satisfying (2.1) and its properties have been well established; see Stroock and Varadhan (1979) and Øksendal (2000). Multivariate diffusion processes have been increasingly useful for stochastic modeling due to their ability to model inter-dependence among components of X_t. However, the transitional densities for multivariate processes generally do not have closed form expressions except for some special processes, as is the case for univariate diffusion processes. Research in Duffie, Pan and Singleton (2000) shows that the CCF is available for multivariate affine processes which are the mostly employed processes in applications. A diffusion process is called affine if $\mu(X_t;\theta)$ and $\sigma(X_t;\theta)\sigma(X_t;\theta)^T$ are linear functions of the state variable.

A distinct feature of diffusion processes is the continuity of their sample paths. Empirical observations collected in financial applications often exhibit jumps, due to arrivals of surprising news to the market. To accommodate this discontinuous feature, a jump diffusion process has been considered by adding a jump component to (2.1):

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dB_t + dZ_{t;\theta}, \tag{2.2}$$

where $Z_{t;\theta}$ is a pure jump process with the occasions of jumps determined by an intensity function $\lambda(X_t)$ and the size of jumps follows a distribution ν. The presence of the jumps makes the processes (2.2) even less likely to admit a closed form expression for the transitional density than the diffusion process.

A special case of (2.2) is the following univariate Ornstein-Uhlenbeck process with compound Poisson jump, which is also called Vasicek with Merton Jump model (VSK-MJ) in finance (Merton, 1976). The model is

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t + J_t dN_t, \tag{2.3}$$

where κ, α and σ are unknown parameters and represent the mean reverting rate, long-run mean and volatility of the process respectively, N_t is a Poisson process with intensity λ, and J_t is the random jump size independent of the filtration $\mathcal{F}_t$ up to time t and has a normal density $N(0, \eta^2)$. The jump component in (2.3) is a compound Poisson process whose Lévy measure is integrable over $R_0 = R \backslash \{0\}$ (R less zero), and hence (2.3) is called finite-activity jump process. The transitional density of model (2.3) is unknown, while its CCF is

$$\psi_t(u;\theta) = exp\{\frac{\sigma^2 u^2}{4\kappa}(e^{-2\kappa\delta} - 1) - \lambda\delta + \gamma + i(\alpha u(1 - e^{-\kappa\delta}) + ue^{-\kappa\delta}X_t)\}, \tag{2.4}$$

where $\gamma = \lambda/(2\kappa)\int_{e^{-2\kappa\delta}}^{1} exp(-\eta^2 u^2 y/2)/ydy$, by solving some ordinary differential equations (Duffie, Pan and Singleton, 2000).

The latest extensions of processes that incorporate jumps include Lévy driven processes with the so-called infinity-activity jumps whose Lévy measure is not integrable over R_0. Although the theory of Lévy processes are well established as documented in Sato (1999), properties of the Lévy driven processes are still being developed and this development forms an exciting part of stochastic analysis nowadays. Recently, Barndorff-Nielsen and Shephard (2001) have proposed the following Lévy driven Ornstein-Uhlenbeck process for stochastic volatility modeling:

$$dX_t = -\lambda X_t dt + dL_{\lambda t}, \qquad X_0 > 0. \tag{2.5}$$

Here L_t is a Lévy process with no Brownian part, a nonnegative drift and a Lévy measure which is zero on the negative half line, and the parameter λ is positive. Since L_t is an increasing process and $X_0 > 0$, the process X_t is strictly stationary over the positive half-line. If X_0 is distributed as Inverse Gaussian with parameters a and b, denoted as $IG(a, b)$, then the marginal distribution of X_t is distributed as $IG(a, b)$ with the mean a/b and the variance a^3/b. This process is called Inverse Gaussian Orstein-Uhlenbeck (IG-OU) process in which L_t jumps infinitely often in every finite time interval, therefore it is an infinite-activity jump process. Barndorff-Nielsen and Shephard (2001) show that $X_t = e^{-\lambda t} X_0 + e^{-\lambda t} \int_0^{\lambda t} e^s dL_s$. Based on this fact and the relationship between the cumulant function of $IG(a, b)$ and the cumulant function of L_t at time $t = 1$, the CCF can be derived as

$$\psi_t(u;\theta) = exp\{-a(\sqrt{-2iu + b^2} - \sqrt{-2iue^{-\lambda\delta} + b^2}) + iue^{-\lambda\delta} X_t\}. \tag{2.6}$$

Here the transitional density is not known in closed form, nor is its approximation so far. This is a good example to highlight the need for estimation and testing methods based on the CCF.

3 Parameter Estimation

An underlying assumption in this section is that the transitional density of the process is not analytically available, but $\psi_t(u;\theta)$ is. Otherwise there is no need to attempt the CCF based estimation as the maximum likelihood estimation is asymptotically efficient under certain conditions.

Let $\{X_{t\delta}\}_{t=1}^n$ be n discretely sampled observations of a d-variate parametric Markov process. For notation simplification, we denote $X_{t\delta}$ as X_t. The sampling interval δ is an arbitrary fixed quantity corresponding to either high or low sampling frequency. We also use $\bar{a}$ and $A^\star$ to denote the conjugate of the complex number a and the conjugate transpose of the complex matrix A, respectively.

Let $\epsilon_t(\tau;\theta) = w(u, r; X_t)\{e^{iu^T X_{t+1}} - \psi_t(u;\theta)\}$ for $\tau = (u^T, r^T)^T \in R^{2d}$, where $w(u, r; X_t)$ is a weight factor. Here $\epsilon_t(\tau;\theta)$ can be regarded as "residuals" between $e^{iu^T X_{t+1}}$ and the parametric CCF $\psi_t(u;\theta)$. The complex weight factor $w(u, r; X_t)$ satisfies $\bar{w}(u, r; X_t) = w(-u, -r; X_t)$ and $|w(u, r; X_t)| = 1$ for any $u, r \in R^d$. The use of a weight factor is designed to utilize more model information.

Let θ_0 be the true parameter value, which is the unique solution of

$$E\{e^{iu^T X_{t+1}} - \psi_t(u;\theta)|X_t\} = 0 \quad \text{for all } u \in R^d. \tag{3.1}$$

Assume that $\psi_t(u;\theta)$ has continuous second derivatives with respect to θ in a neighborhood of θ_0.

Due to the Markov property and (3.1), for any $\tau = (u^T, r^T)^T \in R^{2d}$

$$E\{\epsilon_t(\tau;\theta_0)\} = 0 \quad \text{and} \quad Cov\{\epsilon_{t_1}(\tau;\theta_0), \epsilon_{t_2}(\tau;\theta_0)\} = 0 \quad \text{if} \quad t_1 \neq t_2. \tag{3.2}$$

Hence, $\{\epsilon_t(\tau;\theta_0)\}_{t=1}^n$ constitutes a sequence of martingale differences.

Let $\epsilon_t^R(\tau;\theta)$ and $\epsilon_t^I(\tau;\theta)$ be the real and imaginary parts of $\epsilon_t(\tau;\theta)$ respectively, and $\vec{\epsilon}_t(\tau;\theta) = \left(\epsilon_t^R(\tau;\theta), \epsilon_t^I(\tau;\theta)\right)^T$ be the real bivariate vector corresponding to $\epsilon_t(\tau;\theta)$.

We now formulate an empirical likelihood for θ based on the parametric CCF $\psi_t(u;\theta)$. The empirical likelihood (EL) introduced in Owen (1988) is a technique that allows construction of a nonparametric likelihood for parameters of interest. Despite that the EL method is intrinsically nonparametric, it possesses two important properties of a parametric likelihood, the Wilks' theorem and the Bartlett correction; see Owen (2001) for overviews on empirical likelihood method.

Let $p_1(\tau), \cdots, p_n(\tau)$ be probability weights allocated to the "residuals" $\{\vec{\epsilon}_t(\tau;\theta)\}_{t=1}^n$. An empirical likelihood for θ at τ is

$$L_n(\tau,\theta) = \max \prod_{t=1}^n p_t(\tau) \tag{3.3}$$

subject to $\sum_{t=1}^n p_t(\tau) = 1$ and $\sum_{t=1}^n p_t(\tau)\vec{\epsilon}_t(\tau;\theta) = 0$. Here the second constraint reflects (3.1). The maximum empirical likelihood is attained at $p_t(\tau) \equiv n^{-1}$ for all t such that the maximum likelihood $L_n(\tau;\theta) = n^{-n}$. Let $\ell_n(\tau;\theta) = -2\log\{L_n(\tau;\theta)/n^{-n}\}$ to be the local log-empirical likelihood ratio of θ at τ.

Employing the empirical likelihood algorithm (Owen, 1988), the optimal $p_t(\tau)$ of the above optimization problem (3.3) is

$$p_t(\tau) = \frac{1}{n} \frac{1}{1 + \lambda(\tau;\theta)^T \vec{\epsilon}_t(\tau;\theta)},$$

where $\lambda(\tau;\theta)$ is a Lagrange multiplier in R^2 that satisfies

$$Q_{1n}(\tau;\theta,\lambda) =: \frac{1}{n} \sum_{t=1}^{n} \frac{\vec{\epsilon}_t(\tau;\theta)}{1 + \lambda(\tau;\theta)^T \vec{\epsilon}_t(\tau;\theta)} = 0. \tag{3.4}$$

Hence, the local empirical likelihood ratio becomes

$$\ell_n(\tau;\theta) = 2 \sum_{t=1}^{n} \log\{1 + \lambda(\tau;\theta)^T \vec{\epsilon}_t(\tau;\theta)\}. \tag{3.5}$$

By integrating $\ell_n(\tau;\theta)$ against a probability weight $\pi(\tau)$ which is supported on a compact set S in R^{2d}, an integrated empirical likelihood ratio for θ is

$$\ell_n(\theta) = \int_{\tau \in R^{2d}} \ell_n(\tau;\theta)\pi(\tau) d\tau. \tag{3.6}$$

The maximum empirical likelihood estimator (MELE) for θ is defined as

$$\hat{\theta}_n = \arg\min_{\theta} \ell_n(\theta),$$

by noting that -2 has been multiplied in the EL ratio $\ell_n(\tau;\theta)$.

By combining results in Kitamura (1997) and Chen, Härdle and Li (2003) for the empirical likelihood of α-mixing processes, which are readily satisfied due to the property (3.2), we obtain that

$$\lambda(\tau;\theta) = A_n^{-1}(\tau;\theta) n^{-1} \sum_{t=1}^{n} \vec{\epsilon}_t(\tau;\theta) + \tilde{o}_p(1), \tag{3.7}$$

where $A_n(\tau;\theta) = n^{-1} \sum_{t=1}^{n} \vec{\epsilon}_t(\tau;\theta)\vec{\epsilon}_t(\tau;\theta)^T$ and $\tilde{o}_p(1)$ means $o_p(1)$ uniformly for $\tau \in S$, and

$$\ell_n(\tau;\theta) = \sum_{t=1}^{n} \vec{\epsilon}_t^{\,T}(\tau;\theta) A_n^{-1}(\tau;\theta) n^{-1} \sum_{t=1}^{n} \vec{\epsilon}_t(\tau;\theta) + \tilde{o}_p(1). \tag{3.8}$$

It is more convenient to express $\ell_n(\tau;\theta)$ in complex numbers. Define

$$M_0 = \tfrac{1}{2} \begin{pmatrix} 1 & 1 \\ i^{-1} & -i^{-1} \end{pmatrix} \quad \text{and} \quad \tilde{\epsilon}_t(\tau;\theta) = (\epsilon_t(\tau;\theta), \epsilon_t(-\tau;\theta))^T.$$

Then, $\vec{\epsilon}_t(\tau;\theta) = M_0 \tilde{\epsilon}_t(\tau;\theta)$ and

$$\ell_n(\theta) = n \int \tilde{\epsilon}^{\star}_{(n)}(\tau;\theta) A^{-1}(\tau;\theta_0;\theta) \tilde{\epsilon}_{(n)}(\tau;\theta)\pi(\tau) d\tau \{1 + o_p(1)\} \tag{3.9}$$

where $\tilde{\epsilon}_{(n)}(\tau;\theta) = n^{-1}\sum_{t=1}^{n} \tilde{\epsilon}_t(\tau;\theta)$ and $A(\tau;\theta_0,\theta) = Var\{\tilde{\epsilon}_1(\tau;\theta)\}$.

Like Qin and Lawless (1994), we first show that there exists a consistent estimator $\hat{\theta}_n$ with a certain rate of convergence as follows.

Lemma 1. Under conditions C1-C4 given in the appendix, with probability one, $\ell_n(\theta)$ attains its minimum at $\hat{\theta}_n$ in the interior of the ball $||\theta - \theta_0|| \leq O(n^{-1/3})$, and $\hat{\theta}_n$ and $\lambda(\tau;\hat{\theta}_n)$ satisfy

$$\begin{cases} Q_{1n}(\tau;\hat{\theta}_n, \lambda(\tau;\hat{\theta}_n)) = 0 \quad \text{for all} \quad \tau \in S \quad \text{and} \\ \int Q_{2n}(\tau;\hat{\theta}_n, \lambda(\tau;\hat{\theta}_n))\pi(\tau)\, d\tau = 0, \end{cases} \tag{3.10}$$

where Q_{1n} is defined in (3.4) and

$$Q_{2n}(\tau;\theta,\lambda) = \frac{1}{n}\sum_{t=1}^{n} \frac{1}{1+\lambda(\tau;\theta)^T \vec{\epsilon}(\tau;\theta)} \frac{\partial \vec{\epsilon}^T(\tau;\theta)}{\partial\theta}\lambda. \tag{3.11}$$

We introduce some notations before we derive the asymptotic normality of the $\hat{\theta}_n$ given in Lemma 1. Let $A(\tau_1,\tau_2;\theta_0,\theta) = Cov\{\tilde{\epsilon}_1(\tau_1;\theta), \tilde{\epsilon}_1(\tau_2;\theta)\}$,

$$\Gamma(\theta_0) =: \int E\left(\frac{\partial \tilde{\epsilon}^{\star}_1(\tau;\theta_0)}{\partial\theta}\right) A^{-1}(\tau,\tau;\theta_0,\theta_0) E\left(\frac{\partial \tilde{\epsilon}_1(\tau;\theta_0)}{\partial\theta}\right)\pi(\tau)d\tau, \tag{3.12}$$

and

$$\begin{aligned} V(\theta_0) \quad &= \quad \int\int E\left(\tfrac{\partial \tilde{\epsilon}^{\star}_1(\tau_1;\theta_0)}{\partial\theta}\right) A^{-1}(\tau_1,\tau_1;\theta_0,\theta_0) A(\tau_1,\tau_2;\theta_0,\theta_0) \\ &\quad \times A^{\star -1}(\tau_2,\tau_2;\theta_0,\theta_0) E\left(\tfrac{\partial \tilde{\epsilon}_1(\tau_2;\theta_0)}{\partial\theta}\right)\pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2. \end{aligned} \tag{3.13}$$

Theorem 1: Under Conditions C1-C4 given in the appendix, for the estimator $\hat{\theta}_n$ in Lemma 1, we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,\Sigma)$$

where $\Sigma = \Gamma^{-1}(\theta_0)V(\theta_0)\Gamma^{-1}(\theta_0)$.

The proposed estimator attains the $\sqrt{n}$-rate of convergence. It is computationally stable because computing $\ell_n(\tau;\theta)$ for one τ at a time is essentially one dimensional problem.

Potentially, to make the estimation more efficient, we could formulate the EL estimator simultaneously over a set of $\tau_1, \cdots, \tau_J$, namely let

$$L_{n,r}(\theta) = \max \prod_{t=1}^{n} p_t$$

subject to $\sum_{t=1}^{n} p_t = 1$ and $\sum_{t=1}^{n} p_t \vec{\epsilon}_t(\tau_j;\theta) = 0$ for $j = 1, \cdots, J$. For better efficiency, J should be made large and in fact when $J \to \infty$, $\{\tau_j\}_{j=1}^{J}$ forms a dense set in R^{2d}. This becomes a high dimensional inference problem and the computation is much more involved. In particular, when J is large, the covariance matrix of $(\vec{\epsilon}_t^T(\tau_1;\theta), \cdots, \vec{\epsilon}_t^T(\tau_J;\theta))^T$ may not be invertible.

Carrasco, Chernov, Florens and Ghysels (2007) consider CCF based generalized method of moment estimation by considering a continuum of τ's in a functional space via covariance operator. Again, the covariance operator may not be invertible due to zero eigenvalues. Ridging is needed to avoid the issue, which makes the computation quite involved. Our proposal is effectively a marginal averaging algorithm that utilizes all the model information over a wide range of frequency domain with less computational burden. At the same time, it has a reasonable level of efficiecy. Indeed, simulation studies reported in Section 5 showed that the proposed estimator had efficiency close to that of the MLEs.

4 Test for Model Specification

In this section we consider testing for the validity of (1.1) based on the observed sample path via testing for the parametric specification of the CCF $\psi_t(u;\theta)$. Tests for model specification of a continuous-time Markov process have been proposed by Aït-Sahalia, Fan and Peng (2006) and Chen, Gao and Tang (2008). Despite parameter estimation based on the transitional density is asymptotically efficient, it is unclear if a test based on the transitional density is more powerful than one based on the CCF for the hypotheses we will consider later. Of course, the choice is clear when the transitional density does not admit a closed form while the CCF does, since the latter is a test valid at any level of δ.

Let the underlying process that generates the discretely observed sample path $\{X_t\}_{t=1}^n$ be

$$dX_t = \mu(X_t)dt + \sigma(X_t)dL_t, \tag{4.1}$$

whose CCF is $\psi(u; X_t)$. The process (1.1) is a parametric specification of (4.1). To emphasize the dependence of the CCF on X_t, we write in this section $\psi_t(u)$ as $\psi(u, X_t)$, $\psi_t(u;\theta)$ as $\psi(u, X_t; \theta)$ and other quantities in a similar fashion.

We consider testing

$$H_0 : \psi_t(u) = \psi_t(u; \theta_0) \quad \text{for all } u \in R^d \text{ and some } \theta_0 \in \Theta \text{ almost surely,}$$

against a sequence of local alternative hypotheses

$$H_1 : P\{\psi_t(u) = \psi_t(u; \theta_0) + c_n\Delta_n(u; X_t) \quad \text{for all } u \in R^d\} = 1,$$

where $\{c_n\}$ is a sequence of non-random real constants converging to zero at a certain rate, and $\{\Delta_n(u; X_t)\}$ is a sequence of bounded complex functions which are continuous at $u = 0$ and $\Delta_n(0; X_t) \equiv 0$. See Condition C6 in the appendix for extra restriction on $\Delta_n(u; x)$.

One could use $\ell_n(\hat{\theta}_n)$ as the test statistic where $\ell_n(\theta)$ is the empirical likelihood ratio used in Section 2 to derive the proposed estimator. This seems to be attractive since it is available as a by-product of the estimation. However, the test based on such a test statistic would have a power with uncertain properties although it allows a smaller order of c_n than the test proposed shortly. To illustrate this point, we note from (3.9) and an expansion of $\hat{\theta}_n$ given in (A.7) that

$$\begin{aligned} \ell_n(\hat{\theta}_n) &= n \textstyle\int \tilde{\epsilon}^{\star}_{(n)}(\tau; \theta_0) A^{-1}(\tau, \tau; \theta_0, \theta_0)\tilde{\epsilon}_{(n)}(\tau; \theta_0)\pi_1(\tau)d\tau \\ &\quad -2nS^{\star}_n(\theta_0)\Gamma^{-1}(\theta_0)S_n(\theta_0) + o_p(n^{-1/2}) \end{aligned} \tag{4.2}$$

for both H_0 and H_1, where

$$S_n(\theta_0) = \int \frac{\partial \tilde{\epsilon}^{\star}_{(n)}(\tau;\theta_0)}{\partial\theta} A^{-1}(\tau,\tau;\theta_0,\theta_0)\tilde{\epsilon}_{(n)}(\tau;\theta_0)\pi(\tau)d\tau.$$

The second term on the right hand side of (4.2) is due to estimating θ with $\hat{\theta}_n$, and it is of the same order as the first term on the right hand side. The minus sign in the second term looks rather discouraging for the power since it reduces the mean of the first term by noting that the means of both terms in (4.2) are positive.

To obtain a test with a more predicable power, we propose a kernel smoothed version of $\ell_n(\theta)$. Generally speaking, smoothing is needed when the target of inference is a conditional quantity like the conditional mean in regression. We have avoided smoothing in the parameter estimation due to $\{\epsilon_t(\tau;\theta_0)\}_{t=1}^n$ being martingale differences.

Let K be a kernel function which is a symmetric probability density in R^d, and h be a smoothing bandwidth that tends to 0 as $n \to \infty$. A smoothed version of $L_n(\tau,\theta)$ is

$$L_{nh}(\tau,x;\theta) = \max \prod_{t=1}^{n} p_t(\tau,x), \tag{4.3}$$

subject to $\sum_{t=1}^n p_t(\tau,x) = 1$ and $\sum_{t=1}^n p_t(\tau,x)K_h(x-X_t)\vec{\epsilon}(\tau,X_t;\theta) = 0$.

Let $\ell_{nh}(\tau,x,\theta) = -2\log\{L_{nh}(\tau,x,\theta)n^n\}$ be the log-EL ratio. Then, the integrated log-EL ratio for θ near $X_t = x$ is

$$\ell_{nh}(\theta) = \int\int \ell_{nh}(\tau,x,\theta)\pi_1(\tau)\pi_2(x)d\tau dx,$$

where π_1 and π_2 are probability weight functions on the frequency space and the state space respectively. We can choose π_1 to be the same as π in Section 2.

The test statistic is $\ell_{nh}(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the empirical likelihood estimator. As a matter of fact, we can employ any estimator with $n^{1/2}$-rate of convergence. To appreciate the meaning of the test statistic, let $W_h(x-X_t) = K_h(x-X_t)/\sum_{j=1}^n K_h(x-X_j)$ be the Nadaraya-Watson kernel weight, $\epsilon_{n,h}(\tau,x;\theta) = \sum_{t=1}^n W_h(x-X_t)\epsilon(\tau,X_t;\theta)$ be the kernel smooth of the residuals, $\tilde{\epsilon}_{n,h}(\tau,x;\theta) = (\epsilon_{nh}(\tau,x;\theta), \epsilon_{nh}(-\tau,x;\theta))^T$, and $R(K) = \int K^2(t)dt$. It can be shown by a similar derivation in Chen, Härdle and Li (2003) that

$$\begin{aligned} \ell_{nh}(\theta) &= nh^d R^{-1}(K) \int\int \tilde{\epsilon}^{\star}_{n,h}(\tau,x;\theta) V^{-1}(\tau,x;\theta_0,\theta)\tilde{\epsilon}_{n,h}(\tau,x;\theta) \\ &\quad \times \pi_1(\tau) f(x)\pi_2(x) d\tau dx + O_p\{(nh^d)^{-1/2}\log^3(n) + h^2\log^2(n)\}, \end{aligned} \tag{4.4}$$

where $V(\tau,x;\theta_0,\theta) = Var\{\tilde{\epsilon}(\tau,X_t;\theta)|X_t = x\}$ and $f(x)$ is the density of X_t. So, the test statistic is asymptotic equivalent to a L_2-measure of the averaged "residuals" $\tilde{\epsilon}^{\star}_{n,h}(\tau,x;\theta)$ inversely weighted by the covariance matrix function V. Hence, the proposed test is similar in tune to the tests of Fan and Zhang (2003) for testing diffusion processes, and of Härdle and Mammen (1993) and Wang and Van Keilegom (2007) for testing regression functions.

Let $\ell_{nh,1}(\theta_0)$ be the first term on the right hand side of (4.4) with $\theta = \theta_0$. As shown in (A.12) in the appendix,

$$\ell_{nh}(\hat{\theta}_n) = \ell_{nh,1}(\theta_0) + O_p\{(nh^d)^{-1/2}\log^3(n) + h^2\log^2(n) + h^d\}. \tag{4.5}$$

Hence, unlike the unsmoothed test statistic $\ell_n(\hat{\theta})$, the effects of parameter estimation does not contribute to the leading order term in $\ell_{nh}(\hat{\theta}_n)$.

To describe the power property, we choose $c_n = n^{-1/2}h^{d/4}$ and define

$$\eta_n(\tau,X_t) = w(\tau;X_t)\Delta_n(u,X_t), \qquad \tilde{\eta}_n(\tau,X_t) = (\eta_n(\tau,X_t), \eta_n(-\tau,X_t))^T,$$

$$\mu_n = \int\int \tilde{\eta}^{\star}_n(\tau,x)V^{-1}(\tau,x;\theta_0,\theta_0)\tilde{\eta}_n(\tau,x)\pi_1(\tau)\pi_2(x)f(x)d\tau dx,$$

$$\text{and } \sigma_n^2 = 2R^{-2}(K)h^{-d}\gamma^2(K,V,\pi_1,\pi_2),$$

where $\gamma^2(K,V,\pi_1,\pi_2)$ is defined in (A.18) in the appendix.

The asymptotic normality of $\ell_{nh}(\hat{\theta}_n)$ is shown in the following theorem.

Theorem 2 Under Conditions C1-C6 given in the appendix,

$$h^{-d/2}(\ell_{nh}(\hat{\theta}_n) - 2 - h^{d/2}\mu_n) \xrightarrow{d} N(0, 2R^{-2}(K)\gamma^2(K,V,\pi_1,\pi_2)). \tag{4.6}$$

We note that $\mu_n = 2$ under H_0. Under H_1, since $\Delta_n(u, x)$ is non-vanishing with respect to u, $\tilde{\eta}_n(\tau, x)$ is non-vanishing with respect to u for all x in the support of f, which leads to a positive quantity μ_n due to $V^{-1}(\tau, x; \theta_0, \theta_0)$ being a Hermitian matrix. Since no restriction has been imposed on the functional form of $\Delta_n(u, X_t)$, it means that the test is powerful for a wide range of local alternatives.

Indeed, if $\hat{\gamma}^2(K, V, \pi_1, \pi_2)$ is a consistent estimator of $\gamma^2(K, V, \pi_1, \pi_2)$, the asymptotic normality based test for H_0 with α-level of significance rejects H_0 if

$$\ell_{nh}(\hat{\theta}_n) \geq 2 + z_{1-\alpha}\sqrt{2}h^{d/2}R^{-1}(K)\hat{\gamma}(K, V, \pi_1, \pi_2),$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. Theorem 2 implies that the power of the test under H_1 is

$$\Phi\left(-z_{1-\alpha} + \frac{R(K)\mu_n}{\sqrt{2}\gamma(K, V, \pi_1, \pi_2)}\right),$$

where Φ is the standard normal distribution function.

To make the test less sensitive to the choice of smoothing bandwidth, we propose carrying out the test based on a set of bandwidths, say $\{h_1, \cdots, h_k\}$, for a fixed integer k. This means that we have a set of the EL ratio
$\{\ell_{nh_1}(\hat{\theta}_n), \cdots, \ell_{nh_k}(\hat{\theta}_n)\}$ and the overall test statistic is

$$T_n = \max_{1\leq i\leq k}\{h_i^{-d/2}(\ell_{nh_i}(\hat{\theta}_n) - 2)\}. \tag{4.7}$$

We propose the following bootstrap procedure to find the critical value of the test to reduce finite sample error in approximating the nominal level of significance. The convergence in Theorem 2 and the mapping theorem are the keys that justify the bootstrap based test.

Let t_α be the $1-\alpha$ quantile of T_n where $\alpha \in (0, 1)$ is the nominal size of the test. The following parametric bootstrap procedure is employed to approximate t_α:

Step 1: Simulate a sample path $\{X_t^*\}_{t=1}^n$ at the same frequency δ according to the model under H_0 with the CCF based estimate $\hat{\theta}_n$.

Step 2: Let $\tilde{\theta}_n^*$ be the estimate of θ under H_0 using the resample path $\{X_t^*\}_{t=1}^n$ obtained in Step 1, and T_n^* be the version of T_n for the resampled path.

Step 3: For a large positive integer B, repeat Steps 1 and 2 B times and obtain, after ranking, $T_n^{(1)*} \leq T_n^{(2)*} \leq \cdots \leq T_n^{(B)*}$.

Then, the Monte Carlo approximation of t_α is $T_n^{([B(1-\alpha)]+1)*}$. The proposed test rejects H_0 if $T_n(\hat{\theta}_n) \geq T_n^{([B(1-\alpha)]+1)*}$.

5 Simulation Studies

5.1 Simulation Models

We report in this section the results from our simulation studies which are designed to verify the proposed parameter estimator and model testing procedure. To evaluate the quality of the proposed EL estimator, we first chose two univariate diffusion processes with known transitional densities, so that the MLEs can be compared with the proposed EL estimates. The two processes are the Vasicek model (Vasicek, 1977) (VSK),

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{5.1}$$

and the Cox-Ingersoll-Ross Model (Cox, Ingersoll and Ross, 1985) (CIR),

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t, \tag{5.2}$$

where κ, α and σ are unknown parameters which represent the mean reverting rate, long-run mean and volatility of the process respectively. Both processes are widely used in interest rate modeling and various option price formulation. For the Vasicek model, the transitional distribution of $X_{t+1}|X_t$ is a normal distribution $N(\alpha + (X_t - \alpha)exp(-\kappa\delta), \sigma^2(1 - exp(-2\kappa\delta))/(2\kappa))$. For the CIR model, when $2\kappa\alpha/\sigma^2 > 1$, $X_{t+1}|X_t$ is a multiple of a non-central Chi-square random variable with degrees of freedom $4\kappa\alpha/\sigma^2$ and non-centrality parameter $cX_t exp(-\kappa\delta)$, where the multiplier is $1/c$ and $c = 4\kappa/(\sigma^2(1-exp(-\kappa\delta)))$. The CCFs of these two models can easily be derived from the known transitional densities.

We then considered estimation for the jump diffusion model VSK-MJ as given in (2.3) based on its CCF function (2.4). For comparison, we approximated its transitional density by a mixture of normal distributions, $(1-\lambda\delta)N(\mu_\delta, \sigma_\delta^2) + \lambda\delta N(\mu_\delta, \sigma_\delta^2 + \eta^2)$, which is a first order approximation proposed in Aït-Sahalia, Fan and Peng (2006). Here, $\mu_\delta = \alpha + (X_t - \alpha)exp(-\kappa\delta)$, and $\sigma_\delta^2 = \sigma^2(1 - exp(-2\kappa\delta))/(2\kappa)$. The approximate MLEs were obtained based on the mixture approximation given above.

The Inverse Gaussian OU process in (2.5) whose CCF is specified in (2.6) was also included in the simulation study. Since neither the exact transitional density nor its approximation is available, we were content with carrying out estimation with the proposed methods.

The last simulation model considered for the estimation is a bivariate extension of the univariate Ornstein-Uhlenbeck process, which we call bivariate OU process (BI-OU),

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{5.3}$$

where $X_t = (X_{1t}, X_{2t})$,

$$\kappa = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}, \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \text{and } \sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}.$$

It is the most basic bivariate model capturing mean reversion in the process. Under the condition that the eigenvalues of the matrix κ have positive real parts, the process is stationary with transitional distribution being a bivariate normal $N(m(\delta, X_t), \Omega(\delta))$ where $m(\delta, X_t) = \alpha + exp(-\kappa\delta)(X_t - \alpha)$, $\Omega(\delta) = \Sigma - exp(-\kappa\delta)\Sigma exp(-\kappa^T\delta)$ and

$$\Sigma = \frac{1}{2tr(\kappa)Det(\kappa)}\{Det(\kappa)\sigma\sigma^T + \{\kappa - tr(\kappa)\}\sigma\sigma^T\{\kappa - tr(\kappa)\}^T\}.$$

Here $exp(-\kappa\delta)$ is the matrix exponential. The CCF of the process is known to be $\psi_t(u_1, u_2; \theta) = exp\{iu^T m(\delta, X_t) - u^T\Omega(\delta)u/2\}$ for $u = (u_1, u_2)^T$.

We then carried out simulations to evaluate the ability of the proposed tests in detecting model deviations. When we chose the simulation models, we had in mind two issues in finance that have drawn considerable research attention recently. The first issue is whether the process is subject to jumps, and the second is whether we could differentiate two processes with different jump rates. Our simulation study formulated two settings of hypotheses to address these two issues. In the first setting, we tested

$$H_0 : \text{The process is the VSK model.}$$

In the second setting, we tested

$$H_0 : \text{The process is the jump diffusion model VSK-MJ.}$$

For computing the powers, in the first setting we used the data simulated from H_1 : the jump diffusion model VSK-MJ to test the null model which does not have jumps; in the second setting, we used the data simulated from H_1 : the Inverse Gaussian OU model which has infinite-activity jumps to test the null hypothesis that prescribes a finite-activity jump process.

5.2 Simulation Implementation

For each model, we simulated 500 sample paths which were observed at monthly observations ($\delta = 1/12$) for $n = 125, 250, 500$ respectively. For the VSK, CIR and bivariate OU processes, since the transitional and stationary distributions are known in closed form, we generated the simulated sample paths using the exact transitional densities with the initial value X_0 generated from the stationary distributions. For the jump diffusion model VSK-MJ, since Poisson processes can be well approximated by Bernoulli processes if δ is small, we simulated sample paths by the following discretization,

$$X_{(t+1)\delta_1} = \alpha + (X_{t\delta_1} - \alpha)exp(-\kappa\delta_1) + \sigma\sqrt{\frac{1 - exp(-2\kappa\delta_1)}{2\kappa}}Z + \omega_t, \tag{5.4}$$

where Z is a standard $N(0,1)$, $\omega_t = N_t J_t$, $P(N_t = 1) = \lambda\delta_1 exp(-\lambda\delta_1)$, and $J_t \sim N(0, \eta^2)$. Here the discretization was made at a much smaller sampling interval $\delta_1 = \delta/200$, and we kept every $200 - th$ observation out of entire sequence simulated, leaving the data at the monthly frequency. For the Inverse Gaussian OU process, we referred to Schoutens (2003, p. 69) for simulating the Inverse Gaussian OU process and also used a similar arrangement for saving the data by discarding 199 out of 200 simulated draws.

In parameter estimation, we discovered that for both real and imaginary parts of the CCF, their nonparametric smoothing estimators are wave-like functions and roughly diminish to zero at the same points, which creates a region denoted as S_t (here the subscript t indicates that the region depends on X_t). In practice, we searched on a couple of grid points in the data range of X_t and picked the union of S_t as the support region S for the frequency domain of $\psi_t(u;\theta)$ in the estimation. We then chose the weight function π as a uniform density over the support region.

In model testing, similar effort was initially made to obtain the support region of the nonparametric CCF estimate, denoted as S_{NP}, and the support region of the theoretical CCF under H_0, denoted as S_{H_0}. Here the theoretical CCF under H_0 used $\hat{\theta}_n$ from our EL method. Then the support region of the frequency domain in testing was taken as the union of S_{NP} and S_{H_0}. We chose π_1 to be the uniform density over this support region for testing. We believed that there would be little contribution to the integrated empirical likelihood ratio $\ell_{nh}(\hat{\theta}_n)$ from outside the support region.

The biweight Kernel $K(u) = 15/16(1-u^2)^2 I(|u| \leq 1)$ was used for smoothing in testing. The bandwidth selection was guided by the Cross-Validation (CV) (Härdle, 1990). We pre-ran the simulated 500 sample paths and obtained the average and standard deviation of the bandwidths prescribed by the CV. The bandwidth set contains 5 bandwidths consisting of the average CV bandwidth, one and two standard deviations below or above the average CV bandwidth. The bandwidth sets were specified in Table 3 and 4 for the two test settings. It is observed that the values of the bandwidths were quite small, which was due to the rapid oscillation of the CCF curves which favored smaller bandwidth in the curve fitting. We chose $w(u, r; X_t) = e^{irX_t}$ throughout our simulation study as encouraged by the results in Carrasco, Chernov, Florens and Ghysels (2007).

5.3 Simulation Results

Table 1 reports the empirical averages of the parameter estimates and their standard errors as well as the true parameter values used for simulation. When the sample size increases, standard errors of the proposed all estimates decrease, indicating the consistency of the estimators. We observe from Table 1 (a)-(b) for the VSK and CIR models where the MLEs are available, the proposed EL estimates are quite close to the MLEs. Although the EL estimates tend to have larger standard errors than the MLEs, we do note that under the VSK model in Table 1 (a), the bias of EL estimates for the mean reverting parameter κ are smaller than the corresponding MLEs for all $n = 125$, $n = 250$ and $n = 500$. For the jump diffusion model VSK-MJ (Table 1 (c)), we see the EL estimates are consistently more efficient than the approximate MLEs in the estimation of κ and the Poisson intensity λ. For the Inverse Gaussian OU model which does not have the MLE to compare with, the proposed estimates as reported in Table 1 (d) are close to the true values and the standard errors converge as the sample size increases.

Table 2 reports the estimates for the bivariate OU process and shows that the EL estimates are close to the corresponding MLEs, providing the further evidence of the effectiveness of our EL estimator for multivariate process estimation. We also found that the EL estimates for the long run mean α_1 and the volatility σ_{11} of the first process have smaller biases and standard errors than the MLEs for all $n = 125$, $n = 250$ and $n = 500$.

Tables 3 and 4 report the empirical size and power of the proposed test based on $B = 250$ bootstrap resampled paths for each simulation. They contain the sizes and powers for the overall test that is based on the five bandwidth set, and for the tests that only use one bandwidth. We observe that the tests gave satisfactory sizes under both testing settings. In the first test where we used the data from the jump diffusion model VSK-MJ to test the continuous diffusion model VSK, the powers range from 65% to 95% across the different sample sizes and bandwidths. In the second test where we used data simulated from the infinity-activity jump process (the Inverse Gaussian OU) to test the finitie-activity jump process (the jump diffusion VSK-MJ), the powers range from 71% to 90% across the different sample sizes and bandwidth choices.

6 A Case Study

In this section, we examine empirically the capability of our testing procedure in detecting jumps using the secondary market quotes of the 3-month Treasury Bill (T-bill) between January 1, 1965 and February 2, 1999. This bill was sampled at monthly frequency, and in total we had 410 observations. We used monthly data because our simulation set-up was based on monthly frequency in order to speed up the computation. The data set covers the same time period as the one in Johannes (2004) except that he used daily data. The mean of these bills is 0.065, the volatility is 0.026, the mean of the differences is very close to zero (1.5×10^{-5}) and the standard deviation of the differences is 0.005. Figure 1 plots the level and difference of this data. The sample period contains some large movements that turn out to coincide with arrivals of macroeconomic news (Johannes, 2004). The goal of this empirical study was to test whether the underlying process is subject to jumps or not.

The proposed parameter estimates under each of the four univariate models considered in the simulation study are reported in Table 5. For comparison, the MLEs or the approximate MLEs are also reported except for the Inverse Gaussian OU model. For the univariate diffusion models VSK and CIR, and the jump diffusion model VSK-MJ, the proposed parameter estimates based on CCF are very similar to the MLEs or the approximate MLEs. The EL estimates of the long-run mean α are 0.059 for VSK and 0.064 for CIR, both of which are close to the summary statistic of mean rates (0.065). In VSK, the average volatility of 3-month T-bill monthly return (difference) is estimated to be $\sigma\sqrt{\delta} = 0.018\sqrt{1/12} = 0.005$, which is also close to the summary statistic of volatility for the change (0.005). However the conditional volatility of monthly change in CIR model is $\sigma\sqrt{\delta X_t}$, and X_t has a long-run average 0.064 which is less than 1. Therefore, the process needs to have higher σ (0.057) to bring up the average volatility of monthly change to the same level reflected by the real data. In the jump diffusion model VSK-MJ, our estimate of λ suggests on average about 2 jumps per year. Relative to VSK and CIR models, the estimate for parameter σ in the jump diffusion VSK-MJ model is much smaller (0.008), indicating that allowing jumps in the process helps capturing large movements in the interest rate, and as a result the continuous part of the process does not have to be as volatile as the one in VSK or CIR models. The estimated mean of marginal distribution of the Inverse Gaussian model is $a/b = 1.139/12.558 = 0.091$, which has a reasonable magnitude by referring Figure 1.

We then applied the proposed test for the validity of each of the four models. The bandwidth prescribed by the CV was 0.01. By exploring the kernel estimators of the CCF, a reasonable range for h was from 0.01 to 0.018, that offered smoothness from slightly undersmoothing to slightly oversmoothing. The bandwidth range used in our empirical study consisted of five equally spaced bandwidths ranging from 0.01 to 0.018. Table 6 reports p-values of single bandwidth and the overall tests for the four models. There is no empirical support for VSK model. CIR model performs a little bit better as the distances between the test statistics and the critical values decrease, but the model is still rejected at significance level of 0.05 in the overall test and almost all the single bandwidth tests.

We can not reject the jump diffusion model VSK-MJ in the overall test and the single bandwidth tests except the one with the smallest bandwidth (p-value = 0.046). This constitutes a strong evidence for the presence of jumps and implies that adding (finite-activity) jumps does help capturing the underlying dynamics of the interest rates. By allowing the infinite-activity jumps in the models, the p-values of the tests for the Inverse Gaussian OU model are very supportive even for the small bandwidths, suggesting that the infinite-activity jump model might potentially model the dynamics of the 3-month T-bill rates better.

Our empirical analysis suggests the presence of the jumps in the 3-month T-bill rates. This finding is consistent with the analysis in Johannes (2004) which revealed that "jumps play a dominant role in interest rate dynamics". Our analysis also suggests that the Inverse Gaussian OU process might perform better than the jump diffusion model VSK-MJ in modeling the 3-month T-bill rates. A possible reason for it is that the jump diffusion model VSK-MJ can only generate small continuous movements from Brownian motion and big spikes from the compound Poisson component, but it could miss the movements that are between (i.e. the movements with median sizes). However, the Inverse Gaussian OU process is more flexible since it can generate small, median, and big movements with infinite arrival rates, therefore it could fill in a gap in the jump diffusion model VSK-MJ by capturing movements that are too large for Brownian motion to model but too small for the compound Poisson process to capture.

7 Conclusion

Markov processes encompass a rich collection of stochastic models, from discrete-time to continuous-time models, and from models with continuous sample paths (diffusion) to models that allow jumps. Estimation and testing for Markov processes are challenging statistical inference problems. The challenges arise from several aspects. One is the lack of analytical expression for the transitional density function. Another is, in the context of continuous-time model, that the data are only discretely observed. The sampling interval of the discrete data are critically important for any inference method that replies on approximations for the continuous-time models. We have proposed in this paper a marginal integrated empirical likelihood approach for parameter estimation and model specification test based on the conditional characteristic functions as they are more likely to be available. The approach has robust theoretical and empirical performance for a wide range of processes including processes with jump components.

Appendix

The following conditions are required in our analysis.

C1: The stochastic processes given in (4.1) and (1.1) admit a unique weak solution which is strictly stationary.

C2: (Smoothness) $\psi_t(\tau;\theta) =: \psi(\tau;\theta, X_t)$ and $E\{\epsilon_t(\tau;\theta)\}$ are third continuous differentiable with respect to θ within a neighborhood of θ_0 which is defined in C3. $\pi(\cdot)$ is a bounded probability density supported on a compact set $S \subset R^d$; and the "diffusion" function $\sigma(x)$ is positive definite almost surely.

C3: The parameter space Θ is an open subset of R^p, and the true parameter θ_0 is the unique root of $E\{\epsilon_t(\tau;\theta)\} = 0$ for all $\tau \in S$; and for any $\theta_1 \neq \theta_2$, $P\left\{\psi_t(\cdot;\theta_1) \neq \psi_t(\cdot;\theta_2, X_t)\right\} > 0$.

C4: (Invertibility) The Hermitian matrix $Var\{\tilde{\epsilon}_t(\tau;\theta_0)\}$ is positive definite almost everywhere for $\tau \in R^{2d}$ with respect to the Lebesgue measure in R^{2d}; $E\left(\frac{\partial\tilde{\epsilon}^\star(\tau;\theta_0)}{\partial\theta}\right)$ is of rank 2 (full rank) almost everywhere in R^{2d}; $\Gamma(\theta_0)$ defined in (3.12) is invertible; and the smallest eigenvalue of $E(\frac{\partial\bar{\epsilon}^T(\tau;\theta_0)}{\partial\theta})A^{-1}(\tau,\tau;\theta_0,\theta_0)E(\frac{\partial\bar{\epsilon}(\tau;\theta_0)}{\partial\theta})$ is uniformly bounded away from zero for all τ in S.

C5: The kernel $K(\cdot)$ is a r-th order symmetric kernel supported on $[-1,1]^d$ and has bounded second derivative. We assume $d < 4$ and the smoothing bandwidth $h = O\{n^{-1/(d+2r)}\}$. The bandwidth set $\{h_1,\cdots,h_k\}$ satisfies $h_i = c_i h$ for constants c_i such that $c_1 < c_2 < ... < c_k$ where k is an integer not depending on n.

C6: $\{\Delta_n(u;X_t)\}$ is a sequence of complex functions continuous at $u=0$ and $\Delta_n(0;X_t) \equiv 0$, $\sup_n|\Delta_n(u;X_t)| \leq M_1$ almost surely and the Lebesgue measure of $\{u|\Delta_n(u,x) \neq 0\}$ is positive for all x in the support of the marginal density f, and $c_n = n^{-1/2}h^{-d/4}$ which is the order of the difference between H_0 and H_1.

We need C1 as the basic condition for the stochastic processes involved. C2 consists of smoothness conditions regarding the CCFs and C3 is for identification of parameters. C4 ensures the covariance matrix is invertible which is easier to be justified for our low dimensional formulation of estimation and testing approaches. C5 on the kernel and bandwidth are standard in nonparametric curve estimation. The assumption of $d < 4$ is to make the bias in the kernel estimation a smaller order of $h^{d/2}$ so that the bias is stochastically negligible relative to $\ell_{nh}(\theta_0)$. The kernel method will encounter the curse of dimension when $d \geq 4$. Also, the commonly used processes in finance and other stochastic modeling tend to have dimension less than 4. The bandwidth selected by either cross validation or the plug-in method satisfies the order specified in C5. The first part of C6 regarding $\Delta_n(u; X_t)$ is to qualify $\psi_t(u;\theta)$ under H_1 as a bona fide characteristic function, whereas the part that requires positive measure on the set $\{u|\Delta_n(u,x) \neq 0\}$ is to make H_1 a genuine sequence of alternative hypotheses.

Proof of Lemma 1. By combining results in Kitamura (1997) and Chen, Härdle and Li (2003) for the empirical likelihood of $\alpha-$mixing processes, we can show that

$$\lambda(\tau;\theta) = A_n^{-1}(\tau;\theta)\{\frac{1}{n}\sum_{t=1}^{n}\vec{\epsilon}(\tau;\theta)\} + o(n^{-1/3}) = O(n^{-1/3}) \tag{A.1}$$

almost surely and uniformly in $||\theta-\theta_0|| \leq n^{-1/3}$ and $\tau^T \in S$. Denote $\theta = \theta_0 + un^{-1/3}$. It follows from (A.1) and Taylor expansion that, uniformly in $||u|| = 1$,

$$\begin{aligned}
& \ell_n(\theta) \\
= & \int\{2\textstyle\sum_{t=1}^n \lambda^T(\tau;\theta)\vec{\epsilon}(\tau;\theta) - \sum_{t=1}^n\{\lambda^T(\tau;\theta)\vec{\epsilon}(\tau;\theta)\}^2\}\pi(\tau)\,d\tau + o(n^{1/3}) \\
= & \int n\{\tfrac{1}{n}\textstyle\sum_{t=1}^n \vec{\epsilon}^T(\tau;\theta_0) + \frac{1}{n}\sum_{t=1}^n \frac{\partial\vec{\epsilon}^T(\tau;\theta_0)}{\partial\theta}un^{-1/3}\}A_n^{-1}(\tau;\theta) \\
& \times\{\tfrac{1}{n}\textstyle\sum_{t=1}^n \vec{\epsilon}^T(\tau;\theta_0) + \frac{1}{n}\sum_{t=1}^n \frac{\partial\vec{\epsilon}^T(\tau;\theta_0)}{\partial\theta}un^{-1/3}\}\pi(\tau)\,d\tau + o(n^{1/3}) \qquad \text{(A.2)} \\
= & \int n\{E(\tfrac{\partial\vec{\epsilon}^T(\tau;\theta_0)}{\partial\theta})un^{-1/3}(1+o(1))\}A^{-1}(\tau,\tau;\theta_0,\theta_0) \\
& \times\{E(\tfrac{\partial\vec{\epsilon}(\tau;\theta_0)}{\partial\theta})un^{-1/3}(1+o(1))\}\pi(\tau)\,d\tau + o(n^{1/3}) \\
\geq & \tfrac{1}{2}cn^{1/3}
\end{aligned}$$

almost surely, where $c > 0$ is the smallest eigenvalue of

$$\sup_{\tau\in S} E(\frac{\partial\vec{\epsilon}^T(\tau;\theta_0)}{\partial\theta})A^{-1}(\tau,\tau;\theta_0,\theta_0)E(\frac{\partial\vec{\epsilon}(\tau;\theta_0)}{\partial\theta}).$$

Similarly,

$$\begin{array}{rcl} \ell_n(\theta_0) & = & \int\{\sum_{t=1}^n \vec{\epsilon}^T(\tau;\theta_0)\}A^{-1}(\tau,\tau;\theta_0,\theta_0)\{\frac{1}{n}\sum_{t=1}^n \vec{\epsilon}(\tau;\theta_0)\}\pi(\tau)\,d\tau + o(1) \\ & = & o(n^{1/3}) \end{array} \tag{A.3}$$

almost surely. This together with (A.2) implies that $\ell_n(\theta)$ has a minimum value in the interior of the ball $||\theta-\theta_0|| \le n^{-1/3}$ and this value satisfies $\frac{\partial}{\partial\theta}\ell_n(\theta)=0$, i.e., the second equation in (3.10) by noting (3.4). The first equation follows directly from (3.4).

Proof of Theorem 1. It follows from limit theorems for martingale difference that

$$\left\{ \begin{array}{l} \frac{\partial}{\partial\theta}Q_{1n}(\tau;\theta_0,0) = \frac{1}{n}\sum_{j=1}^n \frac{\partial}{\partial\theta}\vec{\epsilon}(\tau;\theta_0) \xrightarrow{p} M_0 E\{\frac{\partial}{\partial\theta}\tilde{\epsilon}(\tau;\theta_0)\} \\ \frac{\partial}{\partial\lambda^T}Q_{1n}(\tau;\theta_0,0) = -\frac{1}{n}\sum_{j=1}^n \vec{\epsilon}(\tau;\theta_0)\vec{\epsilon}^T(\tau;\theta_0) \xrightarrow{p} -M_0A(\tau,\tau;\theta_0,\theta_0)M_0^\star \\ \frac{\partial}{\partial\theta}Q_{2n}(\tau;\theta_0,0) = 0 \\ \frac{\partial}{\partial\lambda^T}Q_{2n}(\tau;\theta_0,0) = \frac{1}{n}\sum_{j=1}^n \frac{\partial}{\partial\theta}\vec{\epsilon}^T(\tau;\theta_0) \xrightarrow{p} E\{\frac{\partial}{\partial\theta}\tilde{\epsilon}^\star(\tau;\theta_0)\}M_0^\star \end{array} \right. \tag{A.4}$$

uniformly in $\tau^T \in S$. Put $\delta_n = ||\hat{\theta}_n-\theta_0|| + \sup_{\tau^T\in S}||\lambda(\tau;\hat{\theta}_n)||$. Then it follows from Taylor expansion that

$$\begin{array}{rl} 0 & = Q_{1n}(\tau;\hat{\theta}_n,\lambda(\tau;\hat{\theta}_n)) \\ & = Q_{1n}(\tau;\theta_0,0) + \frac{\partial Q_{1n}(\tau;\theta_0,0)}{\partial\theta}(\hat{\theta}_n-\theta_0) + \frac{\partial Q_{1n}(\tau;\theta_0,0)}{\partial\lambda^T}\lambda(\tau;\hat{\theta}_n) + o_p(\delta_n) \end{array} \tag{A.5}$$

uniformly in $\tau^T \in S$, and

$$\begin{array}{rcl} 0 & = & \int Q_{2n}(\tau;\hat{\theta}_n,\lambda(\tau;\hat{\theta}_n))\pi(\tau)\,d\tau \\ & = & \int\{Q_{2n}(\tau;\theta_0,0) + \frac{\partial Q_{2n}(\tau;\theta_0,0)}{\partial\theta}(\hat{\theta}_n-\theta_0) + \frac{\partial Q_{2n}(\tau;\theta_0,0)}{\partial\lambda^T}\lambda(\tau;\hat{\theta}_n)\}\pi(\tau)\,d\tau \\ & & +o_p(\delta_n). \end{array} \tag{A.6}$$

By (A.4) - (A.6), we have

$$\begin{array}{rl} & \hat{\theta}_n-\theta_0 \\ = & -\Gamma^{-1}(\theta_0)\int E\{\frac{\partial}{\partial\theta}\tilde{\epsilon}^\star(\tau;\theta_0)\}A^{-1}(\tau;\theta_0,\theta_0)M_0^{-1}\vec{\epsilon}_{(n)}(\tau;\theta_0)\pi(\tau)\,d\tau + o_p(\delta_n). \end{array} \tag{A.7}$$

Hence the theorem follows from (A.7) and the central limit theorem for Martingale difference.

Derivation of (4.5). Since $\hat{\theta}_n$ is $\sqrt{n}$-consistent to θ_0,

$$\begin{array}{rcl} \ell_{nh}(\hat{\theta}_n) & = & \ell_{nh1}(\theta_0) + nh^dR^{-1}(K)\{(\hat{\theta}-\theta_0)^TS_{n,h}(\theta_0) + S_{n,h}^\star(\theta_0)(\hat{\theta}_n-\theta_0) \\ & & +(\hat{\theta}_n-\theta_0)^T\Gamma_{n,h}(\theta_0)(\hat{\theta}_n-\theta_0)\} + O_p\{(nh^d)^{-1/2}\log^3(n) \\ & & +h^2\log^2(n)\} \end{array} \tag{A.8}$$

where

$$\begin{aligned}\ell_{nh1}(\theta_0) &= nh^d R^{-1}(K) \int\int \tilde{\epsilon}^\star_{n,h}(\tau, X_t; \theta_0) V^{-1}(\tau, x; \theta_0, \theta_0) \\ &\quad \times \tilde{\epsilon}_{n,h}(\tau, x; \theta_0)\pi_1(\tau) f(x) \pi_2(x) d\tau dx,\end{aligned} \tag{A.9}$$

$$\begin{aligned}S_{n,h}(\theta_0) &= \int\int \frac{\partial \tilde{\epsilon}^\star_{n,h}(\tau,x;\theta_0)}{\partial\theta} V^{-1}(\tau, x; \theta_0, \theta_0)\tilde{\epsilon}_{n,h}(\tau, x; \theta_0) \\ &\quad \times \pi_1(\tau)\pi_2(x) f(x) d\tau dx,\end{aligned} \tag{A.10}$$

$$\begin{aligned}\Gamma_{nh}(\theta_0) &= \int\int \frac{\partial \tilde{\epsilon}^\star_{n,h}(\tau,x;\theta_0)}{\partial\theta} V^{-1}(\tau, x; \theta_0, \theta_0)\frac{\partial \tilde{\epsilon}_{n,h}(\tau,x;\theta_0)}{\partial\theta} \\ &\quad \times \pi_1(\tau)\pi_2(x) f(x) d\tau dx.\end{aligned} \tag{A.11}$$

As $S_{n,h}(\theta_0) = O_p(n^{-1/2})$,

$$\ell_{nh}(\hat{\theta}_n) = \ell_{nh,1}(\theta_0) + O_p\{(nh^d)^{-1/2}\log^3(n) + h^2\log^2(n) + h^d\}. \tag{A.12}$$

Proof of Theorem 2.

Recall that $V(\tau_1, \tau_2, x; \theta_0, \theta) = E\{\tilde{\epsilon}(\tau_1, X_t; \theta)\tilde{\epsilon}^\star(\tau_2, X_t; \theta)|X_t = x\}$. And write $V(\tau, x; \theta_0, \theta) = V(\tau, \tau, x; \theta_0, \theta)$. Notice that

$$\begin{aligned}& \ell_{nh,1}(\theta_0) \\ =\ & nh^d R^{-1}(K)\int\int n^{-1}\textstyle\sum_{t_1=1}^n K_h(x - X_{t_1})\{\tilde{\epsilon}^\star(\tau, X_{t_1}) + c_n\tilde{\eta}^\star_n(\tau, X_{t_1})\} \\ & \times V^{-1}(\tau, x; \theta_0, \theta_0) n^{-1}\textstyle\sum_{t_2=1}^n K_h(x - X_{t_2})\{\tilde{\epsilon}(\tau, X_{t_2}) + c_n\tilde{\eta}_n(\tau, X_{t_2})\} \\ & \times \pi_1(\tau)\pi_2(x) f(x) d\tau dx + o_p(h^{d/2}) \\ =\ & R^{-1}(K)\,(H_{n1} + H_{n2} + H_{n3} + H_{n4}) + o_p(h^{d/2}),\end{aligned} \tag{A.13}$$

where, with the choice of $c_n = n^{-1/2}h^{-d/4}$,

$$\begin{aligned}H_{n1} &= n^{-1}h^d \textstyle\sum_{t_1\neq t_2}\int\int K_h(x - X_{t_1})K_h(x - X_{t_2})\tilde{\epsilon}^\star(\tau, X_{t_1})V^{-1}(\tau, x) \\ &\quad \times \tilde{\epsilon}(\tau, X_{t_2})\pi_1(\tau)\pi_2(x) f(x) d\tau dx, \\ H_{n2} &= n^{-1}h^d \textstyle\sum_{t=1}^n \int\int K_h^2(x - X_t)\tilde{\epsilon}^\star(\tau, X_t)V^{-1}(\tau, x)\tilde{\epsilon}(\tau, X_t) \\ &\quad \times \pi_1(\tau)\pi_2(x) f(x) d\tau dx, \\ H_{n3} &= 2n^{1/2}h^{3d/4}\int\int \tilde{\eta}^\star_n(\tau, x)V^{-1}(\tau, x) n^{-1}\textstyle\sum_{t=1}^n K_h(x - X_t)\tilde{\epsilon}(\tau, X_t) \\ &\quad \times \pi_1(\tau)\pi_2(x) f(x) d\tau dx, \\ H_{n4} &= h^{d/2}\int\int \tilde{\eta}^\star_n(\tau, x)V^{-1}(\tau, x)\tilde{\eta}_n(\tau, x)\pi_1(\tau)\pi_2(x) f(x) d\tau dx.\end{aligned} \tag{A.14}$$

We note that $H_{n2} = 2R(K) + o_p(h^d)$ and and the integral in H_{n3} is $O_p(n^{-1/2})$. Hence, $H_{n3} = O_p(n^{3d/4}) = o_p(h^{d/2})$.

Now consider H_{n1}. Clearly, $E(H_{n1}) = 0$ and the double summation in H_{n1} constitutes a generalized U-statistic of order two with the kernel

$$\begin{aligned}\xi_{t_1,t_2} &= \int\int K_h(x - X_{t_1})K_h(x - X_{t_2})\tilde{\epsilon}^{\star}(\tau, X_{t_1})V^{-1}(\tau, x; \theta_0, \theta_0)\tilde{\epsilon}(\tau, X_{t_2}) \\ &\quad \times\pi_1(\tau)\pi_2(x)f(x)d\tau dx.\end{aligned} \tag{A.15}$$

The U-statistic is degenerate due to $\{\tilde{\epsilon}(\tau, X_{t_2})\}$ being martingale differences.

Let $\sigma_n^2 = \sum_{1\le t_1\ne t_2\le n} \sigma_{t_1,t_2}^2$ where $\sigma_{t_1,t_2}^2 = Var(\xi_{t_1,t_2})$. Then, apply the central limit theorem for generalized U-statistics for α-mixing sequences, we have

$$\sigma_n^{-1} \sum_{t_1\ne t_2} \xi_{t_1,t_2} \xrightarrow{d} N(0, 1). \tag{A.16}$$

Furthermore, it can be shown, for instance by following the route of Hjellvik, Yao and Tjostheim (1998), that $\sigma_n^2 = 2n^2\sigma_{n0}^2\{1 + o(1)\}$ where

$\sigma_{n0}^2 = E_{t_1}E_{t_2}(\xi_{t_1,t_2}^2)$. Here E_{t_i} denote marginal expectation with respect to (X_{t_i}, X_{t_i+1}).

To simplify notation, we write $V(\tau_1, \tau_2, x; \theta_0, \theta_0)$ as $V(\tau_1, \tau_2, x)$ and $V(\tau, \tau, x; \theta_0, \theta_0)$ as $V(\tau, x)$. Furthermore, we express the matrices

$$V(\tau_1, \tau_2, x) = (\nu_{lk}(\tau_1, \tau_2, x))_{1\le l,k\le 2} \quad \text{and} \quad V^{-1}(\tau, x) = \left(\nu^{lk}(\tau, x)\right)_{1\le l,k\le 2}.$$

Then, it can be shown that

$$\begin{aligned} & \sigma_{n0}^2 \\ = & \int\int\int\int E_{t_1}E_{t_2}\{K_h(x_1 - X_{t_1})K_h(x_1 - X_{t_2})K_h(x_2 - X_{t_1}) \\ & \times K_h(x_2 - X_{t_2}) \textstyle\sum_{l_1,k_1,l_2,k_2}^2 \epsilon_{l_1}(\tau_1, X_{t_1})\epsilon_{k_1}(\tau_1, X_{t_2})\epsilon_{l_2}(\tau_2, X_{t_1}) \\ & \times\epsilon_{k_2}(\tau_2, X_{t_2})\nu^{l_1,k_1}(\tau_1, x_1)\nu^{l_2,k_2}(\tau_2, x_2)\} \\ & \times\pi_1(\tau_1)\pi_1(\tau_2)f(x_1)f(x_2)\pi_2(x_1)\pi_2(x_2)d\tau_1 d\tau_2 dx_1 dx_2 \\ = & \int\int\int\int E_{t_1}E_{t_2}\{K_h(x_1 - X_{t_1})K_h(x_1 - X_{t_2})K_h(x_2 - X_{t_1}) \\ & \times K_h(x_2 - X_{t_2}) \textstyle\sum_{l_1,k_1,l_2,k_2}^2 V_{l_1l_2}(-\tau_1, \tau_2, X_{t_1})V_{k_1k_2}(\tau_1, -\tau_2, X_{t_2}) \\ & \times\nu^{l_1,k_1}(\tau_1, x_1)\nu^{l_2,k_2}(\tau_2, x_2)\}\pi_1(\tau_1)\pi_1(\tau_2)f(x_1)f(x_2) \\ & \times\pi_2(x_1)\pi_2(x_2)d\tau_1 d\tau_2 dx_1 dx_2 \\ = & h^{-d}\gamma^2(K, V, \pi_1, \pi_2)\{1 + O(h^2)\},\end{aligned} \tag{A.17}$$

where

$$\begin{aligned} & \gamma^2(K, V, \pi_1, \pi_2) \\ = & K^{(4)}(0) \int\int\int \textstyle\sum_{l_1,k_1,l_2,k_2}^2 V_{l_1l_2}(-\tau_1, \tau_2, x)V_{k_1k_2}(\tau_1, -\tau_2, x)\nu^{l_1,k_1}(\tau_1, x) \\ & \times\nu^{l_2,k_2}(\tau_2, x)\pi_1(\tau_1)\pi_1(\tau_2)\pi_2^2(x)f^2(x)d\tau_1 d\tau_2 dx.\end{aligned} \tag{A.18}$$

From (A.16) and (A.17), we have

$$h^{-d/2}H_{n1} \xrightarrow{d} N(0, 2\gamma^2(K, V, \pi_1, \pi_2)) \tag{A.19}$$

This together with the results on H_{n2} and H_{n3} leads to

$$h^{-d/2}(\ell_{nh}(\hat{\theta}) - 2 - \mu_n) \xrightarrow{d} N(0, 2R^{-2}(K)\gamma^2(K, V, \pi_1, \pi_2)) \tag{A.20}$$

where $\mu_n = H_{n4}$. This completes the proof of Theorem 2.

Acknowledgements: The authors acknowledge support from National Science Foundation grants DMS-0604563, DMS-0518904 and SES-0631608.

References

Aït-Sahalia, Y., 2002, Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. Econometrica, **70**, 223 – 262.

Aït-Sahalia, Y., 2008, Closed-form likelihood expansions for multivariate diffusions. Annals of Statistics, **36**, 906 – 937.

Aït-Sahalia, Y., Fan, J. and Peng, H., 2006, Nonparametric transition-based tests for jump-diffusions. Manuscript, Department of Economics, Princeton University.

Aït-Sahalia, Y. and Kimmel, R., 2002, Estimating affine multifactor term structure models using closed-form likelihood expansions. *Manuscript, Department of Economics, Princeton University.*

Aït-Sahalia, Y. and Yu, J., 2006, Saddlepoint approximations for continuous-time markov processes. Journal of Econometrics, **134**, 507 – 551.

Bardorff-Nielsen, O. E., Mikosch, T. and Resnick, S. I., 2001. Lévy Process, Theory and Applications, Birkhäuser, Boston.

Bardorff-Nielsen, O. E. and Shephard, N., 2001, Non-Gaussian Orstein-Uhlenbeck-based models and some of their uses in financial econometrics. *Journal of the Royal*

Statistical Society, Series B, **63**, 167 – 241.

Carrasco, M., Chernov, M., Florens, J-P. and Ghysels, E., 2007, Efficient estimation of jump diffusion and general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, **140**, 529 – 573.

Chacko, G. and Viceira, L. M., 2003, Spectral GMM estimation of continuous-time processes, *J. of Econometrics* **116, no. 1-2**, 259 – 292.

Chen, S. X., Gao, J. and Tang, C.Y., 2008, A test for model specification of diffusion models. Annals of Statistics, **36**, 167 – 198.

Chen, S. X., Härdle, W. and Li, M., 2003, An empirical likelihood goodness–of–fit test for time series. Journal of the Royal Statistical Society, Ser. B., 65, 663 – 678.

Chen, B. and Hong, Y., 2005, Characteristic function-based testing for multifactor continuous-Time markov models via nonparametric regression. Manuscript, Department of Economics, Cornell University.

Chernov, M., Gallant, A., Ghysels, E. and Tauchen, G., 1999, A new class of stochastic volatility models with jumps: theory and estimation. working paper, Columbia University, UNC and Duke University.

Cox, J. C., Ingersoll, J.E., and Ross, S.A., 1985, A theory of term structure of interest Rates. Econometrica, **53**, 385 – 407.

Duffie, D., Pan, J. and Singleton, K., 2000, Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, **68**, 1343 – 1376.

Fan, J., 2005, A selective overview of nonparametric methods in financial econometrics. Statistical Science, **20**, 317 – 357.

Fan, J. and Zhang, C., 2003, A re-examination of diffusion estimators with applications to financial model validation. Journal of Ammerican Statistical Association, **98**, 118 – 134.

Feuerverger, A., 1990, An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics*, **18**, 155 – 161.

Feuerverger, A. and McDunnough, P., 1981, On some Fourier methods for inference. *Journal American Statistical Association*, **76**, 379 – 387.

Feuerverger, A. and *Mureika, R. A.*, 1977, The empirical characteristic function and its applications. *Annal of Statistics* **5**, 88 – 97.

Härdle, W., 1990. Applied Nonparametric Regression, New York: Cambridge University Press.

Härdle, W., and Mammen, E., 1993, Comparing nonparametric versus parametric regression fits. Annals of Statistics, **21**, 1926 – 1947.

Hjellvik, V., Yao, Q. and Tjøstheim, D., 1998, Linearity testing using local polynomial approximation. *J. Statist. Plann. Inference*, **68**, 295 – 321.

Jiang, G. J. and Knight, J. L., 2002, Estimation of continuous-time processes via the empirical characteristic function. Journal of Business and Economic Statistics, **20**, 198 – 212.

Johannes, M., 2004, The statistical and economic role of jumps in continuous-time interest rate models. Journal of Finance, **1**, 227 – 260.

Kitamura, Y., 1997, Empirical likelihood methods with weakly dependent processes. Annals of Statistics, **25**, 2084 – 2102.

Øksendal, B., 2000. Stochastic Differential Equations: An Introduction with Applications, Fifth Edition, Berlin: Springer.

Merton,R., 1976, Option pricing when the underlying stock returns are discontinuous. Journal of Financial Economics, **3**, 125 – 144.

Owen, A., 1988, Empirical likelihood ratio confidence ratios for a single functional. *Biometrika*, **75**, 237 – 249.

Owen, A., 2001, *Empirical Likelihood*, Chapman and Hall.

Qin, J. and Lawless, J., 1994, Empirical likelihood and general estimating functions.

Annals of Statistics, **22**, 300 – 325.

Sato, K., 1999, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.

Schoutens, W., 2003, *Lévy Processes in Finance*, Wiley Series in Probability and Statistics.

Singleton, K. J., 2001, Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, **102**, 111 – 141.

Sørensen, M., 1991, Likelihood methods for diffusions with jumps. In *Statistical Inference in Stochastic Processes,* Ed. by Prabhu, N. U. and Basawa, I. V., 67 - 105.

Strook, D. and Varadhan, S., 1979, *Multivariate Diffusion Processes*, Springer.

Sundaresan, S. M., 2000, Continuous time finance: A review and assessment. *Journal of Finance*, **55**, 1569 – 1622.

Vasicek, O., 1977, An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177 – 188.

Wang, L. and Van Keilegom, I., 2007, Nonparametric test for the form of parametric regression with time series errors. *Statist. Sinica*, **17**, 369 – 386.

Table 1: Empirical averages and their standard errors (in parentheses) of the maximum (MLE) or approximate maximum (AMLE) likelihood estimates and the proposed empirical likelihood estimates (EL) under the four univariate models.

(a) Vasicek Model

n		$\kappa = 0.858$	$\alpha = 0.089$	$\sigma = 0.047$
125	MLE	1.383(0.603)	0.090(0.015)	0.047(0.003)
	EL	1.305(0.643)	0.090(0.017)	0.046(0.004)
250	MLE	1.118(0.397)	0.090(0.011)	0.047(0.002)
	EL	1.052(0.410)	0.089(0.013)	0.046(0.002)
500	MLE	0.966(0.240)	0.089(0.008)	0.047(0.002)
	EL	0.951(0.273)	0.089(0.009)	0.047(0.002)

(b) CIR Model

n		$\kappa = 0.892$	$\alpha = 0.091$	$\sigma = 0.181$
125	MLE	1.372(0.644)	0.091(0.019)	0.183(0.012)
	EL	1.290(0.719)	0.093(0.023)	0.178(0.014)
250	MLE	1.127(0.374)	0.090(0.013)	0.182(0.008)
	EL	1.089(0.435)	0.091(0.015)	0.179(0.009)
500	MLE	1.000(0.245)	0.091(0.010)	0.182(0.006)
	EL	0.977(0.290)	0.092(0.011)	0.180(0.007)

(c) Jump Diffusion VSK-MJ Model

n		$\kappa = 0.858$	$\alpha = 0.089$	$\sigma = 0.047$	$\lambda = 2.0$	$\eta = 0.067$
125	AMLE	1.056(0.381)	0.093(0.020)	0.046(0.005)	1.770(0.723)	0.060(0.016)
	EL	1.090(0.261)	0.084(0.031)	0.048(0.009)	1.851(0.323)	0.066(0.020)
250	AMLE	0.977(0.226)	0.093(0.013)	0.047(0.003)	1.659(0.466)	0.059(0.010)
	EL	1.043(0.201)	0.090(0.023)	0.048(0.007)	1.825(0.236)	0.068(0.015)
500	AMLE	0.939(0.145)	0.092(0.009)	0.047(0.002)	1.620(0.311)	0.060(0.007)
	EL	1.018(0.115)	0.089(0.018)	0.049(0.005)	1.801(0.163)	0.068(0.012)

(d) Inverse Gaussian OU Model

n		$\lambda = 10.0$	$a = 1.0$	$b = 20.0$
125	EL	10.328(3.665)	1.048(0.106)	20.722(2.146)
250	EL	11.154(1.976)	1.059(0.043)	21.380(0.878)
500	EL	11.489(1.652)	1.031(0.024)	20.846(0.461)

Table 2: Empirical averages and their standard errors (in parentheses) of the maximum (MLE) likelihood estimates and the proposed empirical likelihood estimates (EL) under the Bivariate OU model.

n		$\kappa_{11} = 0.22$	$\kappa_{21} = 0.2$	$\kappa_{22} = 0.5$
125	MLE	0.441(0.197)	0.395(0.270)	0.607(0.176)
	EL	0.381(0.208)	0.525(0.238)	0.594(0.192)
250	MLE	0.353(0.165)	0.307(0.148)	0.563(0.110)
	EL	0.354(0.178)	0.449(0.184)	0.564(0.153)
500	MLE	0.280(0.118)	0.241(0.104)	0.526(0.068)
	EL	0.261(0.168)	0.383(0.154)	0.487(0.112)

n		$\alpha_1 = 0.08$	$\alpha_2 = 0.09$	$\sigma_{11} = 0.09$	$\sigma_{22} = 0.17$
125	MLE	0.145(0.166)	0.099(0.056)	0.167(0.067)	0.080(0.079)
	EL	0.141(0.141)	0.117(0.085)	0.129(0.044)	0.071(0.034)
250	MLE	0.141(0.151)	0.096(0.036)	0.140(0.065)	0.116(0.074)
	EL	0.142(0.129)	0.094(0.073)	0.095(0.033)	0.094(0.028)
500	MLE	0.102(0.120)	0.092(0.023)	0.115(0.051)	0.146(0.055)
	EL	0.099(0.108)	0.104(0.064)	0.077(0.024)	0.105(0.028)

Table 3: H_0: VSK versus H_1: the jump diffusion model VSK-MJ

(a) Size Evaluation (in percentage)

n=125	Bandwidth	0.012	0.017	0.021	0.025	0.030	Overall
	Size	4.6	5.6	5.4	5.8	5.6	4.8
n=250	Bandwidth	0.012	0.015	0.018	0.021	0.024	Overall
	Size	5.6	6.2	6.2	6.0	5.8	5.4
n=500	Bandwidth	0.011	0.013	0.015	0.018	0.020	Overall
	Size	5.0	5.6	5.6	5.4	5.6	5.0

(b) Power Evaluation (in percentage)

n=125	Bandwidth	0.016	0.021	0.026	0.032	0.037	Overall
	Power	72.0	71.6	70.4	69.2	65.8	72.2
n=250	Bandwidth	0.016	0.019	0.022	0.026	0.029	Overall
	Power	82.4	82.4	82.2	82.4	82.2	82.6
n=500	Bandwidth	0.014	0.017	0.019	0.021	0.024	Overall
	Power	95.0	94.8	94.6	94.4	94.2	94.8

Table 4: H_0: the jump diffusion model VSK-MJ versus H_1:the Inverse Gaussian OU model

(a) Size Evaluation (in percentage)

n=125	Bandwidth	0.017	0.022	0.028	0.034	0.040	Overall
	Size	3.4	3.6	4.0	3.6	4.6	4.6
n=250	Bandwidth	0.017	0.021	0.024	0.028	0.032	Overall
	Size	4.6	4.6	4.6	4.6	5.0	4.8
n=500	Bandwidth	0.016	0.019	0.021	0.024	0.026	Overall
	Size	5.0	5.2	5.2	5.0	5.0	5.0

(b) Power Evaluation (in percentage)

n=125	Bandwidth	0.008	0.012	0.017	0.021	0.026	Overall
	Power	71.6	73.8	73.2	71.4	71.2	74.4
n=250	Bandwidth	0.008	0.011	0.014	0.017	0.020	Overall
	Power	84.0	84.2	83.4	81.8	81.4	84.4
n=500	Bandwidth	0.008	0.010	0.012	0.014	0.016	Overall
	Power	89.0	88.6	85.6	82.0	79.4	90.6

Table 5: Empirical Estimation for the 3-month T-bill Data

(a) VSK Model

	κ	α	σ
MLE	0.277	0.065	0.019
EL	0.274	0.059	0.018

(b) CIR Model

	κ	α	σ
MLE	0.182	0.066	0.061
EL	0.182	0.064	0.057

(c) VSK-MJ Model

	κ	α	σ	λ	η
AMLE	0.071	0.077	0.009	1.863	0.012
EL	0.072	0.076	0.008	1.862	0.013

(d) Inverse Gaussian OU Model

	λ	a	b
EL	0.264	1.139	12.558

Table 6: P-values for the 3-month T-bill Data

	Bandwidth	0.010	0.012	0.014	0.016	0.018	Overall
VSK	Test Stats	21.971	19.225	16.145	13.267	10.786	14.828
	$l^*_{0.05}$	3.228	3.123	2.845	2.724	2.647	1.462
	P-values	0.0	0.0	0.0	0.0	0.0	0.0
CIR	Test Stats	6.015	4.775	3.755	2.954	2.335	3.546
	$l^*_{0.05}$	2.782	2.739	2.825	2.650	2.448	1.229
	P-values	0.0	0.01	0.02	0.026	0.054	0.0
VSK-MJ	Test Stats	37.204	40.901	45.046	49.878	55.561	25.600
	$l^*_{0.05}$	35.669	43.548	52.247	62.744	74.298	28.751
	P-values	0.046	0.074	0.102	0.126	0.148	0.0880
IG-OU	Test Stats	10.716	9.374	7.962	6.663	5.528	6.870
	$l^*_{0.05}$	40.463	47.665	46.444	42.396	41.750	27.940
	P-values	0.11	0.148	0.124	0.128	0.122	0.162

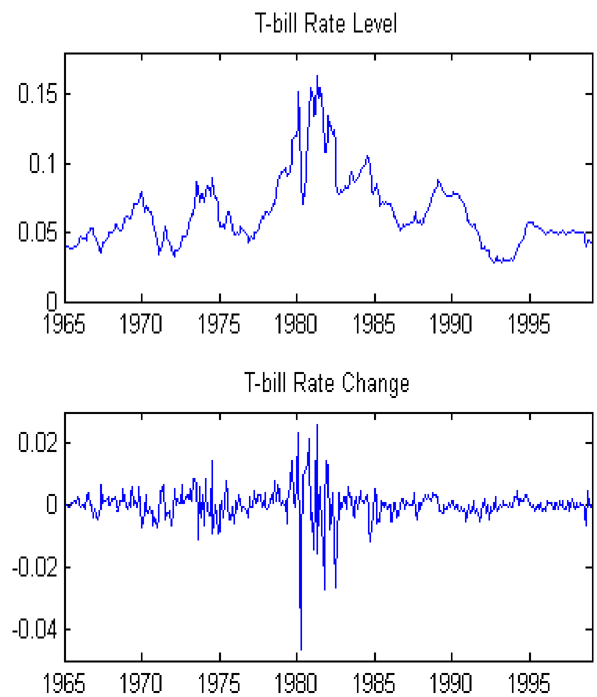

Figure 1: The monthly time series of the level and change in the 3-month T-bill rate, from January 1, 1965 to February 2, 1999